

(51) Int.Cl. ⁷	識別記号	F I	テ-マコ-ト*(参考)
G 0 6 F 17/30	4 1 4 1 1 0	G 0 6 F 17/30	4 1 4 Z 1 1 0 F 1 1 0 E
12/00	5 4 6	12/00	5 4 6 K
13/00	3 5 1	13/00	3 5 1 Z
審査請求 有 請求項の数12 O L (全 13 頁) 最終頁に続く			

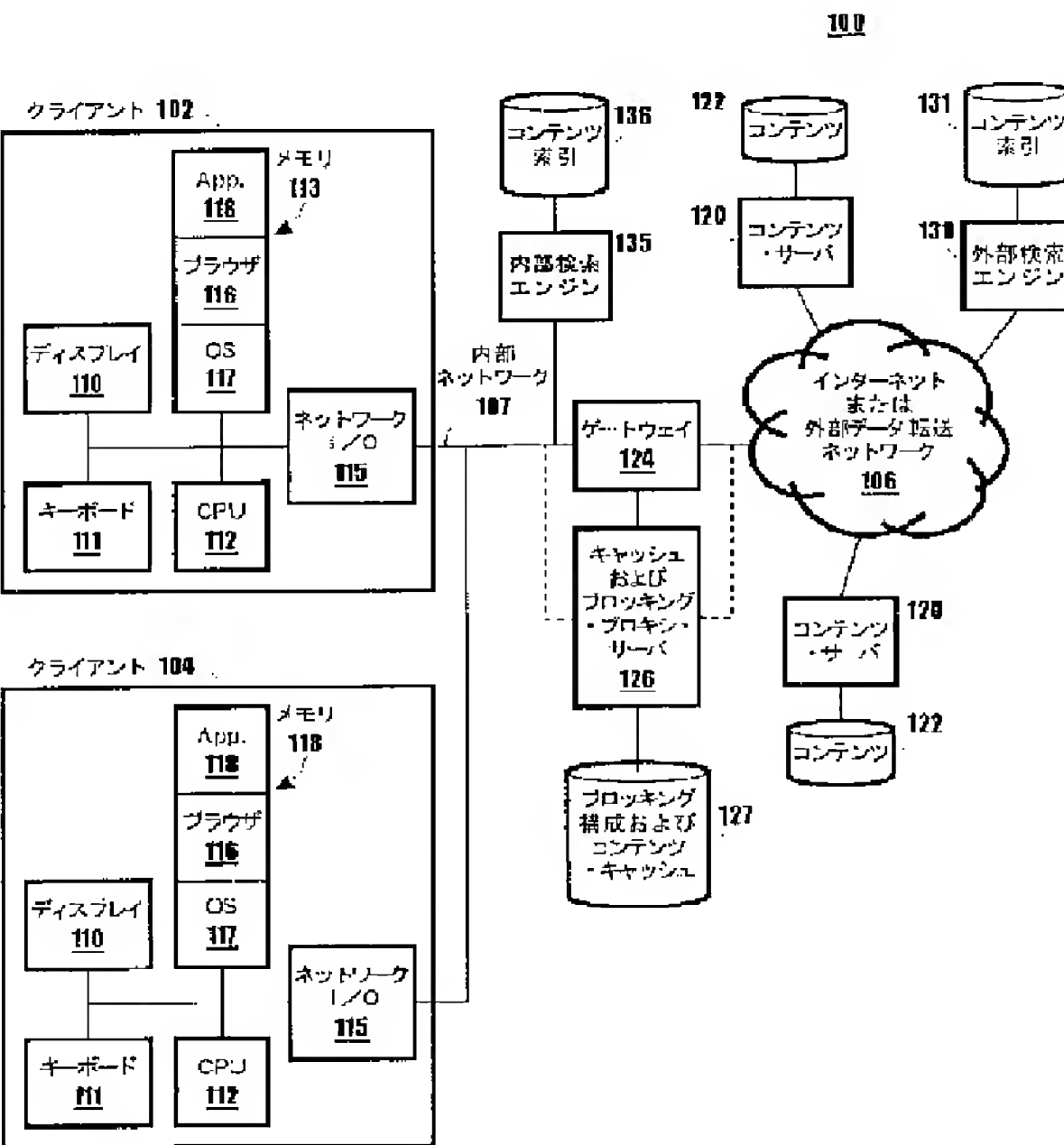
(21) 出願番号	特願2000－121247(P2000－121247)	(71) 出願人	390009531 インターナショナル・ビジネス・マシーンズ・コーポレーション INTERNATIONAL BUSIN ESS MASCHINES CORPO RATION アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
(22) 出願日	平成12年 4 月21日 (2000. 4. 21)	(72) 発明者	レオナルド・シー・マサラーニ アメリカ合衆国06880 コネティカット州、 ウェストポート、 シグナル・レーン 11
(31) 優先権主張番号	0 9 / 3 0 2 8 5 1	(74) 代理人	100086243 弁理士 坂口 博 (外1名)
(32) 優先日	平成11年 4 月30日 (1999. 4. 30)		
(33) 優先権主張国	米国 (U S)		

(54) 【発明の名称】 コンテンツ索引付け検索システム及び検索結果提供方法

(57) 【要約】 (修正有)

【課題】 コンテンツのフィルタリング及びブロッキングを行うコンテンツ索引付け検索システムを提供する。

【解決手段】 ユーザは、プロキシ・サーバとキャッシュ及びブロッキング・エンジンとしての役割を持つゲートウェイを通して検索エンジンに検索問い合わせを行う。ブロッキング・エンジンは、検索結果に関してコンテンツのフィルタリング及びブロッキングを行い、ユーザのコンテンツ検索結果との間の整合性を与える。又、キャッシュ及びフィルタリング・エンジンと同様のコンテンツ・ブロッキング・ポリシーを実行するために、或はキャッシュ及びエンジン・コンテンツを検索することで索引データベースを構築するために、或は検索エンジンがそれ自身の索引付けデータベースを構築するときにキャッシュ及びフィルタリング・エンジンを通して検索エンジンを修正する。



【特許請求の範囲】

【請求項1】コンテンツのフィルタリング及びブロッキング制限と一致する検索結果を提供するコンテンツ索引付け検索システムであって、データベースを含むコンテンツ索引付け検索エンジンと、キャッシュを含むキャッシュ及びブロッキング・プロキシ・サーバと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記コンテンツ索引付け検索エンジンに検索問い合わせを行い、前記キャッシュから検索結果を受け取るための手段と、前記コンテンツ索引付け検索エンジンと結合し、コンテンツのフィルタリング及びブロッキング・ポリシーを実行するブロッキング・エンジンと、前記コンテンツ索引付け検索エンジンを修正して前記ブロッキング・エンジンと同じコンテンツ・ブロッキング・ポリシーを実行するための手段と、を備えるコンテンツ索引付け検索システム。

【請求項2】コンテンツ索引付けフェーズの間、前記コンテンツ索引付け検索エンジンで前記ブロッキング・ポリシーを実行するための手段をさらに備える請求項1に記載のシステム。

【請求項3】エンド・ユーザ検索結果表示フェーズの間、前記ブロッキング・ポリシーを実行するための手段をさらに備える請求項1に記載のシステム。

【請求項4】キャッシュ・コンテンツを索引することによって索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正する手段をさらに備える請求項1に記載のシステム。

【請求項5】前記コンテンツ索引付け検索エンジンが索引付けデータベースを構築するとき前記キャッシュ及びブロッキング・エンジン結果を取り込むために前記コンテンツ索引付け検索エンジンを修正するための手段をさらに備える請求項1に記載のシステム。

【請求項6】データベース及びキャッシュに結合したコンテンツ索引付け検索エンジンと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記キャッシュを介してエンド・ユーザに提供される検索結果に対してコンテンツのフィルタリング及びブロッキング制限を実行するブロッキング・エンジンとを有するコンテンツ索引付け検索システムにおいて、コンテンツのフィルタリング及びブロッキングの実施に一致する検索結果を提供する方法であって、

(a) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、除外パターンに一致する任意の情報サイトURLをスキップさせるステップと、

(b) 前記コンテンツ索引付け検索エンジンのプロセスを変えて、明確に許容される情報サイトURLリストに

一致するサイトまたはルート・コンテンツ・ソースのみを検索するステップと、

(c) 前記キャッシュ及びブロッキング・エンジンで定められるフィルタリングポリシーを前記コンテンツ索引付け検索エンジンで実行するステップとを有し、前記

(C) のステップで前記フィルタリングポリシーは、

(i) 一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンからコンテンツのフィルタリング規則を読み込むステップと、

(i i) 多数索引付けデータベース・ツリーを生成し、各ツリーを前記コンテンツのフィルタリング規則で定められたユーザ・グループと対応付けるステップと、

(i i i) 除外パターンに一致する任意の情報サイト、URL、またはドキュメントのユーザに対しての表示を避けるステップと、

(i v) 明確に許容される情報サイトURLリストに一致するソース由来のドキュメント／コンテンツ・ポインタを表示するステップと、

(v) 前記キャッシュ及び前記ブロッキング・エンジンで定められたフィルタリングプロセスに応じる情報ネットワーク／コンテンツ／ドキュメントのみをユーザに表示するステップとによって定められ、さらに前記(v)のステップで前記フィルタリングプロセスは、

(a a) 一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンから前記コンテンツのフィルタリング規則を読み込むステップと、

(b b) ユーザに対して個々のものまたはグループに対して前記フィルタリング規則によって許容される検索結果のみを表示するステップによって定められる方法。

【請求項7】(d) コンテンツ・エンジン走査標的を情報サイト／URLリストよりはむしろコンテンツ・キャッシュ記憶装置に修正するステップと、

(e) API、データベース走査、及び共有ファイル走査によって前記キャッシュ及びブロッキング・エンジンのURL／コンテンツ／ドキュメント・ツリーをトラバースするステップとをさらに有する請求項6に記載の方法。

【請求項8】(f) エンド・ユーザのブラウザと同様に構成されるように前記検索エンジン・コンテンツ走査及び索引付けプロセスを修正するステップをさらに有する請求項6に記載の方法。

【請求項9】(g) 索引付けデータベースを構築する時に前記キャッシュを通過するように前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する請求項6に記載の方法。

【請求項10】(h) 前記キャッシュを検索することで索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する請求項6に記載の方法。

【請求項11】(i)前記コンテンツ索引付け検索エンジンを内部ネットワークに接続するステップと、(j)前記コンテンツ索引付け検索エンジンを内部ネットワーク操作に接続するステップとをさらに有する請求項6に記載の方法。

【請求項12】前記コンテンツ索引付け検索エンジンを外部ネットワークに接続して組織コンテンツ・ブロッキング・ポリシーに対して整合性を与えるステップをさらに有する請求項6に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、情報検索システムに関する。特に、本発明はブロッキング・エンジンにインプリメントされたコンテンツのフィルタリング及びブロッキングを行うポリシーと一致する検索結果を与えるコンテンツ索引付け検索システム及び方法に関する。

【0002】

【従来の技術】インターネットや他のデータ・ネットワーク及びシステムで利用可能なテキスト及びマルチメディア・コンテンツの爆発的成長によって、エンド・ユーザは、関心ある情報を探し出すためにテキスト及びキー・ワード検索ツールにますます頼っている。エンド・ユーザは、求めている情報及びドキュメントを表しているキー・ワードを検索ツールまたは検索エンジンに入力する。そして、検索ツールまたは検索エンジンは既存の索引付けデータベースで検索を行い、興味あると思われるドキュメントに対するポイントのリストを、ドキュメントのタイトルとともに、また多くの場合ドキュメント本体から抽出されたテキストからなる数行の記述とともに戻ってくる。続いて、エンド・ユーザは検索のために戻ってきたポイントのいくつか、あるいは全てをナビゲートし、実際のドキュメントまたはオンライン上のコンテンツを閲覧する。検索エンジン索引付けデータベースは、一般にコンテンツ・ソース（例えば、インターネット・ウェブ・サイト）に対して自動プログラムを起動し、コンテンツ・ツリー（多くの場合、他のサイトへ移動）へのリンクと同様にルート・コンテンツ・ソースを自動検索し、さらに今後の検索のためにデータベースに含まれる情報を索引付けすることによって、自動的に構築されるか、あるいは半自動的に構築される。インターネット上のウェブ・サイトのような大きいコンテンツ・ソースに対しては、自動化された検索および索引付けは索引検索データベースを作る唯一実用的な方法である。

【0003】オンライン・システム及びネットワーク、企業、個人、グループ、及びネットワーク・サービス・プロバイダ（NSP）に利用可能な情報の多様性が増加することに伴って、エンド・ユーザにとって不適當あるいは不要なものとみなされるコンテンツをふるいにかけるための、さもなければそのようなコンテンツの入手のしやすさを制限するようなポリシーや管理がますます実

施されている。そのようなコンテンツ管理ポリシーは、一般に所定のオンライン・サービス及びネットワークのエンド・ユーザの全てあるいは一部に対して不要なコンテンツが届かないようにブロッキングする。コンテンツのブロッキングは、一般にコンテンツ・プロキシ・ゲートウェイ、データ・ネットワーク・ファイアウォール、あるいはエンド・ユーザと目的とするコンテンツ・ソースとの間に挿入された他のデバイスで実行される。しばしば、コンテンツのフィルタリングはコンテンツ・キャッシュ・エンジンの一部として実現され、ユーザ母集団のために所望のコンテンツのみがキャッシュに保持されて不要のコンテンツがキャッシュされるのを阻止する。すべてのユーザはキャッシュを通してのみネットワーク・コンテンツにアクセスすることができる。コンテンツは、一般にユーザ・グループまたはビジネス用途にとって有害あるいは不適當なものであること、さもなければ一日のうちの特定の時間に見るものである場合にブロッキング・される。しばしば、NSP及び企業は、特定の母集団に関するコンテンツ・サイトまたはドキュメントの適合性を判断するために、評価システムまたはサービス、例えばインターネット・コンテンツ選択のためのプラットフォーム（PICS）に頼るであろう。エンド・ユーザはいくつかのシステムで自分自身で課した一連のブロッキング・ポリシーを選択してもよい。

【0004】自動的に大量のコンテンツを索引付ける自動検索エンジンの必要性和、コンテンツのいくつかがエンド・ユーザに最終的に届くのをブロッキングするブロッキング・エンジンの必要性和との間に位置するNSP及びデータ転送プロバイダに対して、重要な問題が浮上する。特に、問題はフィルタリング及びブロッキング・ポリシーを実施するエンジンと検索エンジンとの一体化及び連携化の欠如である。一体化の欠如はいくつかの理由によって生ずる。すなわち、(a)多くの組織は、検索エンジン、例えばインターネット上で利用可能な検索エンジンに頼る該組織のサイトまたはサービスに対して、コンテンツのフィルタリング及びブロッキング・を行うポリシーを採用して実施している。(b)検索エンジンは意図的に可能な限り多くのコンテンツを検索して索引付けしなければならず、また全てのコンテンツを積極的に求める傾向がある。一方、フィルタリング及びブロッキング・エンジンでは、キャッシュに格納し、かつエンド・ユーザに対して最終的に表示されるドキュメントの選択が意図的に企てられる。

【0005】実行（インプリメンテーション）効率及び高性能に対する要求とともに検索エンジンとブロッキングエンジンとの間に本質的に異なる役割が存在することから、これら二つの情報検索機能の一体化及び連係が妨げられる。

【0006】上記問題は、フィルタリング／ブロッキング・ポリシーにもとづいて最終的にはアクセス不可能

であるコンテンツ・ドキュメントにもかかわらず、検索エンジンのサービスを利用すると同時にそのようなコンテンツ・ドキュメントの記述やタイトルがエンド・ユーザに対して表示されるという事実によって明白となる。首尾一貫性がないことによるエンド・ユーザの不便さやフラストレーションのほかに、検索エンジンによって返されるタイトル及び短い記述は、それら自身が大いに不快であるか、あるいはさもないければ望ましくないものかもしれない。

【0007】したがって、実行可能な程度のわずかなプロトコル及びパフォーマンス効果でパフォーマンス・ブロッキング方式と一致または整合する検索結果が得られる情報検索システムが求められている。

【0008】コンテンツ索引付け検索及びブロッキング・システムとして以下のものが挙げられる。

【0009】1997年12月23日に発行された米国特許第5,701,469号 (Brandliら) は誤って含まれてしまった保存検索結果を結果から取り除き、誤って除外された保存検索結果を加える検索結果収集ルーチンを実行するコンテンツ索引付け検索システムを開示している。このようにして、ユーザの問い合わせに回答して生じた検索結果は、最初の検索結果を生成するのに使われたコンテンツ索引が最新のものでなくても正確に作られる。

【0010】1996年6月27日に出願されて1998年11月10日に発行された米国特許第5,835,722号 (Brandshawら) は、検索不適な材料の生成または転送のためのコンピュータ操作を包括的に監視することによって、不適当な材料の使用及び転送をブロッキングするためのターミナルを開示しており、それによってターミナルブロッキングされるか、あるいは監視介入によってのみブロッキングされる。

【0011】1995年7月5日に出願され1998年1月6日に発行された米国特許第5,706,507号 (Schloss) は、不要な材料をブロッキングまたは感知するために、コンテンツ・サーバからダウンロードされたデータのコンテンツを評価するサード・パーティによって操作されるアドバイザリ・サーバを開示している。

【0012】1997年4月8日に発行された米国特許第5,619,648号 (Canaleら) は、ユーザ対応関係のモデルにもとづいて電子メールのメッセージがユーザに提供されるべきかどうかを判断する電子メール・フィルタリングを開示している。

【0013】いずれの従来技術も、ブロッキング・ポリシーによって許可されたコンテンツのみがコンテンツ検索の結果としてエンド・ユーザに戻されて検索結果はブロッキング・ポリシーに一致したものとなるような、ブロッキング・エンジンに実現されたブロッキング・ポリシーに適合する検索結果を提供するコンテンツ索引付け検索システムを開示していない。

【0014】

【発明が解決しようとする課題】本発明の第1の目的は、検索エンジン結果とコンテンツ・ブロッキング・ポリシーとの間に整合性を与える操作に関する改良された情報検索システム及び方法を提供することである。

【0015】本発明の第2の目的は、ブロッキング・ポリシーと整合する検索結果を与える操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0016】本発明の第3の目的は、キャッシュ及びフィルタリングエンジンでブロッキング・ポリシーを実行する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0017】本発明の第4の目的は、コンテンツ索引付けフェーズの間、ブロッキング・ポリシーを実行する操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0018】本発明の第5の目的は、エンド・ユーザの検索結果を表示するフェーズの間、ブロッキング・ポリシーを実行する操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0019】本発明の第6の目的は、検索の代わりにキャッシュ及びブロッキング・エンジンのローカル・レポシトリを検索し、最終的なコンテンツ・サイト及びコンテンツ・サーバを索引付けする操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0020】本発明の第7の目的は、標的コンテンツに向けてキャッシュ及びフィルタリングエンジンを通過する構成となった操作に関する改良されたコンテンツ索引付け検索システム及び方法を提供することである。

【0021】

【課題を解決するための手段】これらの目的及び他の目的、特徴及び利点は、検索結果が別の実施形態で実行されるエンド・ユーザの組織フィルタリング及びブロッキング・ポリシーと一致するように一般に不要なコンテンツをブロッキングする管理ポリシーを実行するために、検索エンジンとエンド・ユーザとの間を結合するキャッシュ・エンジンとデータベースとを持つコンテンツ索引付け検索エンジンを含む情報検索ネットワークで達成される。

【0022】第1の実施形態例では、ブロッキング・ポリシーに許容されるコンテンツのみが検索エンジン索引付けデータベースに追加される。第2の実施形態例では、検索エンジンの検索及び表示プロセスはブロッキング・ポリシーを実行するために修正される。第3の実施形態例では、検索エンジンの操作及び索引付けオートマトン・プロセスの標的はキャッシュ・エンジンのコンテンツを検索することによって索引付けデータベースを構築するように修正される。第4の実施形態例では、検索エンジンの走査及び索引付けオートマトンはエンド・ユ

ーザのブラウザと同様に構成されるもので、すなわち標的コンテンツに到達するためにキャッシュ及びフィルタリング・エンジンを通過する。

【0023】

【発明の実施の形態】図1において、情報検索システム100は内部または制御されたネットワーク107を介してインターネットまたは他の分散型データ・ネットワーク106と接続している複数のクライアント・デバイス102、104を有する。典型的なクライアントは、ディスプレイ110、キーボード111、CPU112、メモリ113及びネットワーク接続性入出力デバイス115を持つパーソナル・コンピュータ(PC)である。そのようなクライアント及びネットワークの例として、会社内部のネットワークに接続したPCのビジネス・ユーザやサービス・プロバイダのネットワークに接続したPCのホーム・ユーザが挙げられ、どちらの場合も最終的にはより大きなインターネットに接続する。Netscape Communicator, IBM Web Explorer等の商標で販売されているブラウザ116は、標準的なオペレーティング・システム117及びアプリケーション・プログラム118とともにメモリ113にインストールされる。ブラウザ116は、インターネット106につながったコンテンツ・サーバ120からコンテンツの読み込みまたはダウンロードするために、クライアント・デバイス102、104でランまたは実行される。各コンテンツ・サーバは、クライアント102、104等からのコンテンツ要求に対して応答可能なデータを格納するためのデータベース122を有する。一つの形態として、データはテキスト及び他のマルチメディア・コンテンツを含むHTMLドキュメントの集まりとして格納される。

【0024】ゲートウェイ124は、図に示すように、複数のクライアントまたは内部ネットワーク・セグメント107とインターネット106とのインターフェースとして一般に使用される。通常、キャッシュ及びコンテンツのフィルタリング・エンジン126を有するプロキシ・サーバは内部ネットワーク107からインターネット106への接続パスに挿入され、コンテンツ・ブロッキング・ポリシーを実現することでパフォーマンス及び管理を増す。キャッシング及びブロッキング・プロキシ・サーバはゲートウェイに接続することができ、あるいはインターネット107と外部ネットワーク106とに平行して直接的に接続することができる。

【0025】ウェブ・ブラウザ116を実行するクライアント・システム102、104は、HTTP(Hypertext Transfer Protocol)要求を用いてコンテンツ・サーバ120からコンテンツを要求し、HTTP応答でコンテンツを受け取る。HTTP要求及び応答は、クライアントとコンテンツ・サーバとの間の通信リンク上で伝達されるTCP/IPソケット上で生ずる。ユーザは、コンテンツ・サーバに格納されたコンテンツを明確に要

求するか、あるいはコンテンツ・サーバに格納されたコンテンツを指すハイパーリンク・アンカーをとるかによって、コンテンツ要求を生成する。受信すると、ブラウザはHTTPセッションを使用するそのコンテンツをロードする。HTTPに関する詳細な説明は、Berners-Leeらの"Hypertext Transfer Protocol-HTTP/1.0"Draft I EFT-HTTP-V10-Spec-0.0 Text 1995(March 8)に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。HTMLに関する詳細な説明は、Berners-Leeの"Hypertext Markup Language(HTML)"Draft I EFT.IIIR-HTML-01, June 1993(絶版ドラフト)に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。TCP/IPソケットに関する詳細な説明は、W. Richard Stevensの"TCP/IP Illustrated, Vol.1--The Protocols", Addison-Westlake, 1994 pages 1-20, 229-262に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。

【0026】Webブラウザ116を用いるクライアント・システムのユーザは、たいていの場合、キーワード検索手段によってインターネット・コンテンツの場所を突き止めるために限定された従来の検索エンジン・サーバ130、135及びデータベース131、136にそれぞれアクセスする。これらの固定エンジン・サーバは、管理されたネットワーク107に対して外部検索エンジン・サーバ130または内部検索エンジン・サーバ135となる。それらが同一の基本機能を実行する間に、内部に結合した検索エンジン・サーバ135は内部ネットワーク・オペレータによって独立して管理され、どちらかといえば好まれる。この方法の発明は、一般に内部に結合して管理されるサーチ・エンジン135によって、あるいは組織に対するサービスとして首尾一貫して組織にコンテンツ・ブロッキング・ポリシーを与える外部検索エンジン130によって、実現されるであろう。検索エンジン・サーバ130または135に向けられたキーワード検索の結果として、エンド・ユーザは、最終的なコンテンツに対するハイパーリンク・アンカーとして表示されるテキスト抜粋及び一致したURLのリストがWebブラウザ116に表示されるのを見るであろう。ユーザは、Webブラウザ116を用いて一つ以上のコンテンツ一致に対してリンクを選択し、たどることができる。

【0027】図2では、コンテンツ・サンプルをフィルタリング／ブロッキングする構成のテーブル200がクライアントによって、またはネットワーク／サービス・アドミニストレータによって生成され、不適当または不要と見なされるコンテンツの有効性をフィルタリングするか、さもなければ限定するためにプロキシ・サーバ126でインストールされる。これらのコンテンツ・アクセス管理方式は、一般に所定のオンライン・サービスま

たはネットワークでエンド・ユーザの全てまたは一部に対して、望ましくないコンテンツが届くのをブロックする。テーブルはキャッシュ及びフィルタリングエンジン126にインストールされ、通常はデータベース127に格納される。一つの形態として、テーブルは各ユーザまたは各ユーザ・グループに対して、ユーザまたはグループ識別子(ID)203、ブロック・に対するキーワードのリスト205、PICS(Platform For Interconnect Content Selection)規則207、接触すべきではないURLのブラック・リスト209、及び接触することができるURLのみのホワイト・リスト211のうち一つ以上のものを含む行201を有する。URLの説明は、Berners-Leeらの"Uniform Resource Locators(URL)", RFC 1738, December 1994に示されており、本明細書ではこの文献の内容全体を本明細書の一部をなすものとして援用する。PICS評価は、URLを記述するドキュメントに取り込まれたPICラベルにもとづいてURLへのアクセスを許可またはブロックするPICS規則から得られる。PICS規則は、W3C等で発行されたインターネット上の<http://www.w3.org/TR/REC-PICSRules-971229>に記載されている。特に、PICS規則はそれらのURLを記述するPICラベルにもとづいてURLに対するアクセスを許可またはブロックするフィルタリング規則(プロファイル)を表すための言語である。ラベルは、インターネット上の<http://www.w3.org/PICS/>で入手可能なPICS Technicals pecification-1.1にもとづくソフトウェア・ツールを用いることで生成される。ソフトウェア・ツールは、特定のURLを記述するドキュメントでラベルを生成するために使われる。あるいは、ドキュメントの中にラベルを貼る代わりに、他に依存しない読取りプログラムがラベル・ビューロと呼ばれる別のサーバを通してラベルを配る。フィルタリング・ソフトウェアは、消費者が器具または自家用車を論評する特定の雑誌を読むことを知ると同じように、ラベルを見付けるためにラベル・ビューロにおいて調べることを知るのであろう。ひとたびラベルが生成されると、ラベルはWebブラウザに送られたドキュメントのコンテンツに先行するHTTPヘッダ・ストリームに追加のヘッダとして挿入される。あるいは、METAタグを用いてHTMLドキュメントにラベルを埋め込むことができる。この方法で、ラベルは画像、ビデオ、あるいは他のものでもなくHTMLドキュメントだけで送られる。PICS-対応のコンテンツ・サーバがInternational Business Machines Corporation (Armonk, NY)から入手可能である。

【0028】キャッシュ及びブロック・エンジンにインストールされたブロック・テーブルによって、いくつかのプロセス選択肢がコンテンツ・サーチとコンテンツ・ブロック・エンジンとを結合させるために利用可能であり、ブロック・ポリシーによって最終

的に許可されるコンテンツのみがコンテンツ検索の結果としてクライアントに戻される。各個人ユーザに対して異なる規則を持つことが可能であると同時に、全てのユーザに対して適用される規則の単一セットを持つこと、あるいは各々が独自の規則を持つ少数のユーザ・グループにユーザを分けることがよりいっそう処理しやすい。個人ユーザまたはグループは、もし定められるならば、いくつかの手段を使うことによって識別されることが可能である。そのような手段としては、例えばユーザ/グループIDマッピングに対するクライアント・システムIPアドレス、ブラウジング・セッションの開始時にHTTP基本認証を用いること、ユーザ識別を追跡するためのHTTPwebのcookiesを使用することが挙げられる。

【0029】図3に戻ると、プロセス300はコンテンツ索引付けフェーズの間にブロック・ポリシーを実行する。ステップ302では、検索エンジン135によるコンテンツ操作及び索引付けオートマトン・プロセスが修正される。ステップ304ではコンテンツ及びフィルタリングエンジン126からのコンテンツのフィルタリング規則を、通常の間隔で、または変化が検出されるたびに、アプリケーション・プログラム・インタフェース(API)または規則定義ファイルの転送を介して検索エンジン・サーバ135に読み込まれる。ステップ306では、多数の索引付けデータベース・ツリーが必要に応じて生成され、各ツリーはコンテンツのフィルタリング規則で定義されたようにユーザ・グループと対応づけられる。例えば、子供用として厳格なPICフィルタリング規則を付けている一つの索引付けデータベース・ツリーが定義され、大人用としてもっと自由なフィルタリング規則を持つ一つの索引付けデータベース・ツリーが定義される。

【0030】ステップ308では、サーチ・エンジンのオートマトン・プロセスは、コンテンツ・ブロック規則が調べられている間に走査及び索引付けコンテンツの走査及び索引付けを標的サーバのリストから開始する。

【0031】ステップ310では、もしホワイト・リストが存在するならば、検索エンジンは明確に許可可能なサイト/URLリストまたはホワイト・リストに一致するWebサイトまたはルート・コンテンツ・ソースのみを検索する。

【0032】ステップ312では、もし除外すべきURLのブラック・リストが規則の中に設定されるならば、ブラック・リスト・パターンに一致するWebサイトURLはいかなるものでも除外される。

【0033】ステップ314では、索引付けデータベース・ツリーによってサービスされているユーザ・セットに適用されるPICS規則は、処理されているサイト/コンテンツ/ドキュメントに適用され、その結果としてドキュメントが除外されるか、もしくは含まれる。

【0034】ステップ316では、除外するキーワードのリストが特定されるならば、ドキュメント・テキストを走査して一つ以上のキーワードがリストに含まれるならばドキュメントを除去する。

【0035】ステップ318では、そのグループに関してフィルタリング規則によって許容される場合のみ、ドキュメントは適当な索引付けデータベースに加えられる。

【0036】図3のプロセスの利点は、全ての追加の（排除）処理はデータベース索引付けフェーズで実行される。ユーザ検索処理及びプレゼンテーション・フェーズで必要とされる追加の処理はほとんどない。たぶん、可能性のある変化に対してコンテンツが再走査されとしても、検索操作は検索エンジンのライフ・サイクルで索引付け操作よりもかなり頻繁に行われる。

【0037】図4では、エンド・ユーザ検索結果表示フェーズの間に別のプロセス400がブロッキング・ポリシーを実行する。ステップ402では、検索エンジンの走査及び索引付けオートマトン・プロセスは変化せず、単一の索引付けデータベース・ツリーが維持されている。ステップ404では、ブロッキング・ポリシーを適用するためサーチ・エンジンの検索及び表示プロセスが修正される。

【0038】ステップ406では、キャッシュ・エンジンからのコンテンツのフィルタリング規則（API経由または規則定義の転送を介して）は一定間隔で、あるいは変化が検出されるたびに検索エンジンに読み込まれる。

【0039】ステップ407では、ユーザによって始められた検索要求の処理が索引データベースに対して開始される。

【0040】ステップ408では、ユーザ要求を満足させる全ての一致しているドキュメントのリストを作成し、ブロッキング規則の適用のために準備する。

【0041】ステップ410では、明らかに許容されるURLのホワイト・ファイルが規則のなかで特定されるならば、ホワイト・リストに含まれない全ての一致するドキュメントは除外される。

【0042】ステップ412では、除外パターン・リスト（ブラック・リスト）に一致するWebサイト、URL、またはドキュメントのいずれも除外される。

【0043】ステップ414では、もしPICS規則が特定されるならば、PICS規則によって確認されなかったURLはどれでも除外される。

【0044】ステップ416では、もしキーワード・リストが規則に特定されているならば、リストにある一つ以上のキーワードをテキストに含むURLはいずれも除外される。

【0045】ステップ418では、ユーザの要求に一致し、ブロッキング規則を満足するURLポインタの残り

のサブセットがクライアントに対する表示のために戻される。

【0046】図4のプロセスの第1の利点は、索引付けデータベースを再構築する影響なしに最新のポリシーが各検索に適用できることである。単一の索引付けデータベースはすべてのユーザに対して使用することができる。このプロセスによってほとんど影響なしに個々の管理まで下げてフィルタリンググループを変更する定義が可能となる。

【0047】図5では、プロセス500は検索エンジンを修正して、コンテンツをキャッシュしているエンジンのコンテンツを検索することでその索引付けデータベースを構築する。ステップ501では、検索エンジン走査及び索引付けオートマトン・プロセスが修正される。最終的なコンテンツ・ソース・サイトを走査及び索引付けする代わりに、プロセスはキャッシュ及びブロッキング・エンジンのコンテンツのローカル・リポジトリを検索するように設定される。ステップ503では、検索エンジンの走査標的は、サイト／URLリストよりはむしろ適当なコンテンツをキャッシュしている記憶装置に修正される。ステップ505では、キャッシュ及びブロッキング・エンジンのURL／コンテンツ／ドキュメント・ツリーはAPI、データベース操作または共有フィールド・システム操作を介してトラバースされる。ステップ507では、ローカル・インストレーションにおいて一つ以上のユーザ・グループに対するブロッキング・フィルタリング方式に従うので、キャッシュに見いだされる任意のドキュメントが索引付けデータベースに加えられる。

【0048】図5のプロセスの第1の利点はフィルタリング及びブロッキング規則を適用することがそのように設計されたエンジン、すなわちキャッシュ及びブロッキング・エンジンによってたった一回行われることである。走査及び索引付け走査は、より可変的なインターネット・コンテンツ・サイトよりはむしろ標的コンテンツの局所（高性能）コピー上で実行される。

【0049】図6において、検索エンジンはそれ自体の索引付けデータベースを構築するので、プロセス600はキャッシュ及びフィルタリングエンジンを通過するために検索エンジンを修正する。ステップ601では、検索エンジン走査及び索引付けオートマトンを修正してエンド・ユーザのブラウザと同様に構築する。すなわち、標的コンテンツに達するためにHTTPプロキシを用いてキャッシュ及びフィルタリングエンジンを通過する。ステップ603では、検索エンジン・オートマトンは適当なキャッシュ及びフィルタリングエンジンに対して構成されているHTTPプロキシを使用する構成となっている。ステップ605では、コンテンツを走査及び索引付けする一方で、検索エンジン・オートマトンは、ユーザ・グループのポリシーによって許容されるサイト／コ

ンテンツ／ドキュメントを含むサブセットのみをユーザが受け取るように、ユーザ・グループの一つに属するエンド・ユーザをシミュレートする構成となっている。

【0050】図6のプロセスの第1の利点は、検索エンジンを実質的に修正しないことである。コンテンツ・ブロッキング及びフィルタリングは、そのように設計及び最適化されたキャッシュ及びブロッキング・エンジンによって実行される。ブロッキング・ポリシーによって許容されたコンテンツのみが索引付けのために検索エンジンに達する。走査及び索引付けがなされるべきサイト／コンテンツのいくつかはローカル・キャッシュ記憶装置に見いだされるため、検索エンジンの効率及び性能は増加する。

【0051】以上説明したように、ブロッキング・ポリシーによって最終的に許容されるコンテンツに対するポインタのみがコンテンツ検索の結果としてエンド・ユーザに戻るように、コンテンツ検索及びコンテンツ・ブロッキング・エンジンが結合する。コンテンツ検索及びコンテンツ・ブロッキング・エンジンを結合させるために多数のプロセスが記述される。そのようなものとして、本発明はエンド・ユーザのコンテンツ検索の結果と個々の組織のコンテンツのフィルタリング及びブロッキング・ポリシーとの間の整合性を与える。本発明は既存のインターネット及び他のネットワーク、すなわちデータ・プロトコル及び標準に対して、変更の必要性無しにただちに利用することが可能である。

【0052】本発明をインターネット（HTTP／Web）環境について説明したが、同様の概念はデータが検索されるほとんどのデータ及びネットワーク環境に適用される。可能性のある一致のリストは、アクセスまたはコンテンツ管理方式によって許容されるならばデータを順に消費／一覧するエンド・ユーザに対して表示される。特許請求の範囲に定義したように本発明の精神及び範囲から逸脱することなく様々な変更が可能である。

【0053】まとめとして、本発明の構成に関して以下の事項を開示する。

（1）コンテンツのフィルタリング及びブロッキング制限と一致する検索結果を提供するコンテンツ索引付け検索システムであって、データベースを含むコンテンツ索引付け検索エンジンと、キャッシュを含むキャッシュ及びブロッキング・プロキシ・サーバと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記コンテンツ索引付け検索エンジンに検索問い合わせを行い、前記キャッシュから検索結果を受け取るための手段と、前記コンテンツ索引付け検索エンジンと結合し、コンテンツのフィルタリング及びブロッキング・ポリシーを実行するブロッキング・エンジンと、前記コンテンツ索引付け検索エンジンを修正して前記ブロッキング・エンジンと同じコンテンツ・ブロッキング・ポリシーを実行するための手段と、を備えるコンテンツ索引付

け検索システム。

（2）コンテンツ索引付けフェーズの間、前記コンテンツ索引付け検索エンジンで前記ブロッキング・ポリシーを実行するための手段をさらに備える上記（1）に記載のシステム。

（3）エンド・ユーザ検索結果表示フェーズの間、前記ブロッキング・ポリシーを実行するための手段をさらに備える上記（1）に記載のシステム。

（4）キャッシュ・コンテンツを索引することによって索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正する手段をさらに備える上記（1）に記載のシステム。

（5）前記コンテンツ索引付け検索エンジンが索引付けデータベースを構築するときに前記キャッシュ及びブロッキング・エンジン結果を取り込むために前記コンテンツ索引付け検索エンジンを修正するための手段をさらに備える上記（1）に記載のシステム。

（6）データベース及びキャッシュに結合したコンテンツ索引付け検索エンジンと、前記コンテンツ索引付け検索エンジンに結合した情報ネットワークと、前記キャッシュを介してエンド・ユーザに提供される検索結果に対してコンテンツのフィルタリング及びブロッキング制限を実行するブロッキング・エンジンとを有するコンテンツ索引付け検索システムにおいて、コンテンツのフィルタリング及びブロッキングの実施に一致する検索結果を提供する方法であって、（a）前記コンテンツ索引付け検索エンジンのプロセスを変えて、除外パターンに一致する任意の情報サイトURLをスキップさせるステップと、（b）前記コンテンツ索引付け検索エンジンのプロセスを変えて、明確に許容される情報サイトURLリストに一致するサイトまたはルート・コンテンツ・ソースのみを検索するステップと、（c）前記キャッシュ及びブロッキング・エンジンで定められるフィルタリングポリシーを前記コンテンツ索引付け検索エンジンで実行するステップとを有し、前記（c）のステップで前記フィルタリングポリシーは、（i）一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンからコンテンツのフィルタリング規則を読み込むステップと、（ii）多数索引付けデータベース・ツリーを生成し、各ツリーを前記コンテンツのフィルタリング規則で定められたユーザ・グループと対応付けるステップと、（iii）除外パターンに一致する任意の情報サイト、URL、またはドキュメントのユーザに対しての表示を避けるステップと、（iv）明確に許容される情報サイトURLリストに一致するソース由来のドキュメント／コンテンツ・ポインタを表示するステップと、（v）前記キャッシュ及び前記ブロッキング・エンジンで定められたフィルタリングプロセスに応じる情報ネットワーク／コンテンツ／ドキュメントのみをユーザに表示するステップとによって定められ、さらに前

記(v)のステップで前記フィルタリングプロセスは、
 (aa)一定の間隔で、または変化が検知されるたびに、前記キャッシュ及びフィルタリングエンジンから前記コンテンツのフィルタリング規則を読み込むステップと、(bb)ユーザに対して個々のものまたはグループに対して前記フィルタリング規則によって許容される検索結果のみを表示するステップによって定められる方法。

(7)(d)コンテンツ・エンジン走査標的を情報サイト／URLリストよりはむしろコンテンツ・キャッシュ記憶装置に修正するステップと、(e)API、データベース走査、及び共有ファイル走査によって前記キャッシュ及びブロッキング・エンジンのURL／コンテンツ／ドキュメント・ツリーをトラバースするステップとをさらに有する上記(6)に記載の方法。

(8)(f)エンド・ユーザのブラウザと同様にして構成されるように前記検索エンジン・コンテンツ走査及び索引付けプロセスを修正するステップをさらに有する上記(6)に記載の方法。

(9)(g)索引付けデータベースを構築する時に前記キャッシュを通過するように前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する上記(6)に記載の方法。

(10)(h)前記キャッシュを検索することで索引付けデータベースを構築するために前記コンテンツ索引付け検索エンジンを修正するステップをさらに有する上記(6)に記載の方法。

(11)(i)前記コンテンツ索引付け検索エンジンを内部ネットワークに接続するステップと、(j)前記コンテンツ索引付け検索エンジンを内部ネットワーク操作に接続するステップとをさらに有する上記(6)に記載の方法。(12)前記コンテンツ索引付け検索エンジンを外部ネットワークに接続して組織コンテンツ・ブロッキング・ポリシーに対して整合性を与えるステップをさらに有する上記(6)に記載の方法。

【図面の簡単な説明】

【図1】本発明にもとづく情報検索システムの構成を説明するためのブロック図である。

【図2】図1に示す情報検索システムにおいて実行されるコンテンツ・ブロッキング規則の表である。

【図3】コンテンツ検索及び索引付けフェーズの間のブロッキング・ポリシーを実行する第1の実施形態例における図1のシステムの操作を説明するためのフローチャ

ートである。

【図4】エンド・ユーザの検索結果を表示するフェーズの間、ブロッキング・ポリシーを実行する第2の実施形態例における図1のシステムの操作を説明するためのフローチャートである。

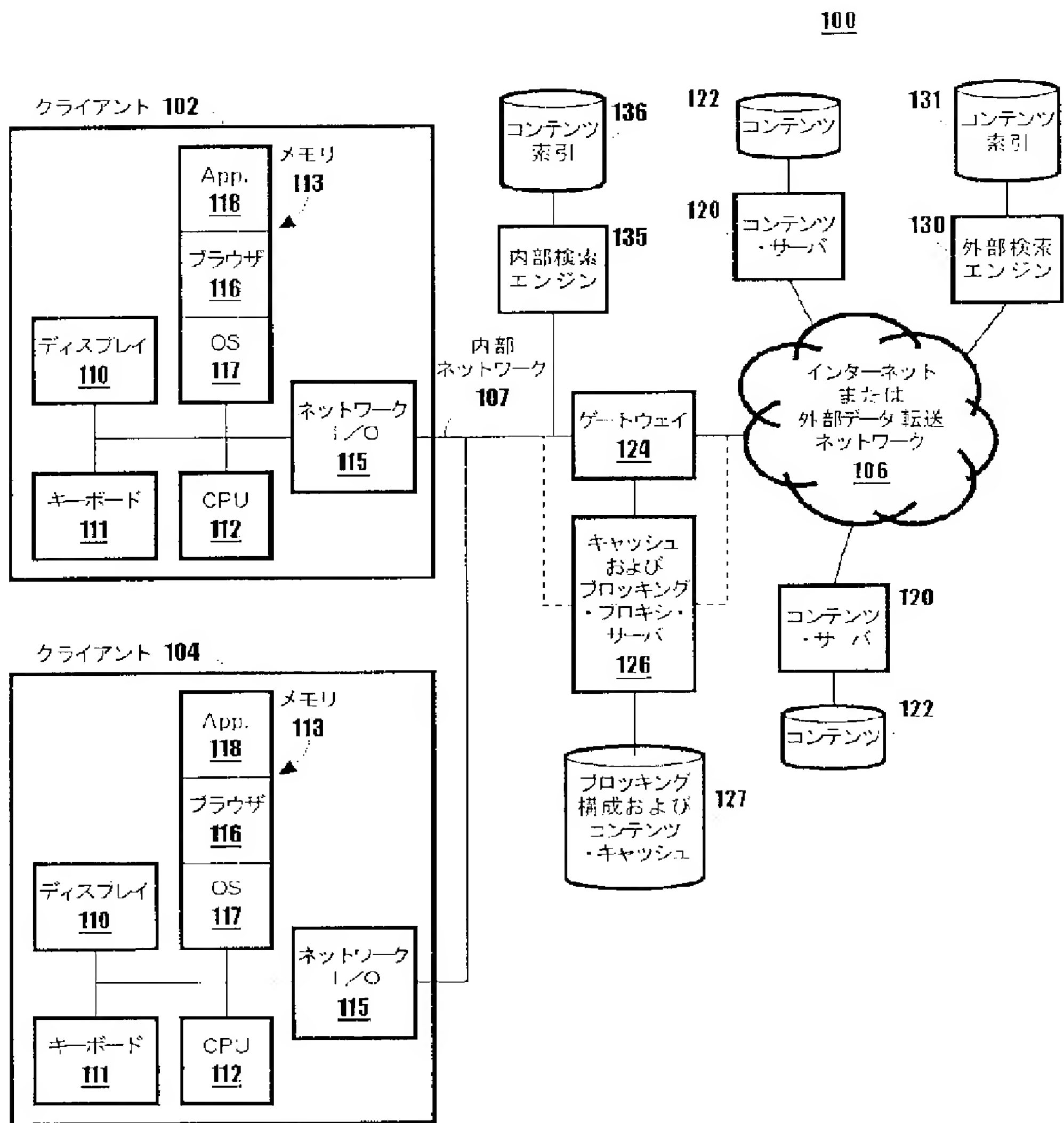
【図5】第3の実施形態例における図1の検索エンジンを説明するためのフローチャートである。

【図6】第4の実施形態例における図1の検索エンジンを説明するためのフローチャートである。

【符号の説明】

100	情報検索システム
102	クライアント・デバイス
104	クライアント・デバイス
106	インターネット
107	ネットワーク
110	ディスプレイ
111	キーボード
112	CPU
113	メモリ
115	ネットワーク接続性入出力デバイス
116	ブラウザ
117	オペレーティング・システム
118	アプリケーション・プログラム
120	コンテンツ・サーバ
122	データベース
124	ゲートウェイ
126	キャッシュ及びコンテンツのフィルタリング・エンジン
130	検索エンジン・サーバ
131	データベース
135	検索エンジン・サーバ
136	データベース
200	テーブル
201	行
203	識別子
204	ユーザまたはグループID
205	ブロッキングに対するキーワードのリスト
207	PICS規則
209	ブラック・リスト
211	ホワイト・リスト
300	プロセス
400	プロセス

【図 1】



【図 2】

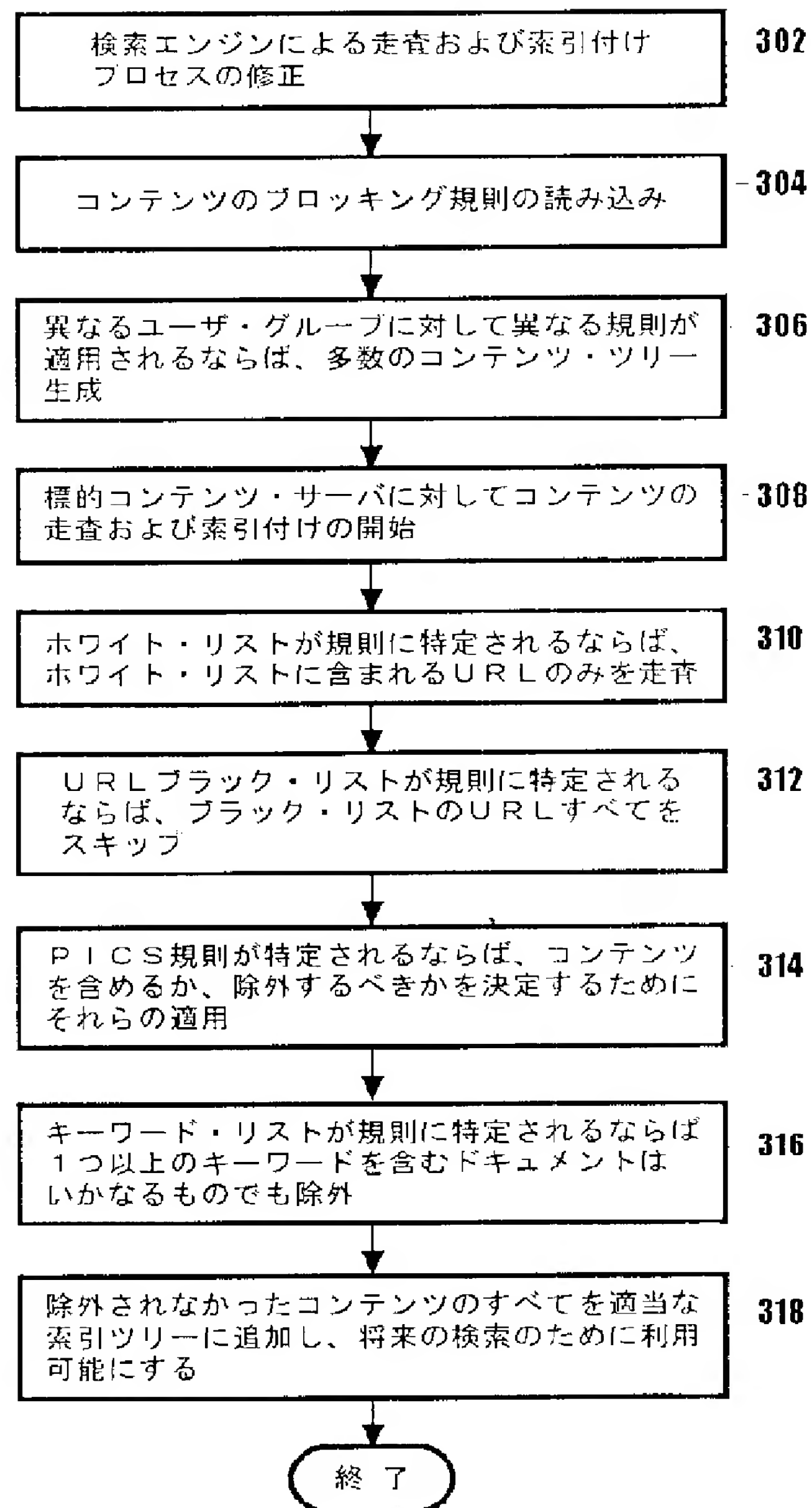
200

ユーザまたはグループ ID 203	ブロックに対するキーワード 205	PICS 規則 207	URL ブラック・リスト 209	URL ホワイト・リスト 211
201 ID	性 酔け事 薬物	ラベル規則 1 ラベル規則 2 ラベル規則 1	URL X URL Y URL Z	URL A URL B URL C

サンプル・コンテンツ・ブロッキング規則表

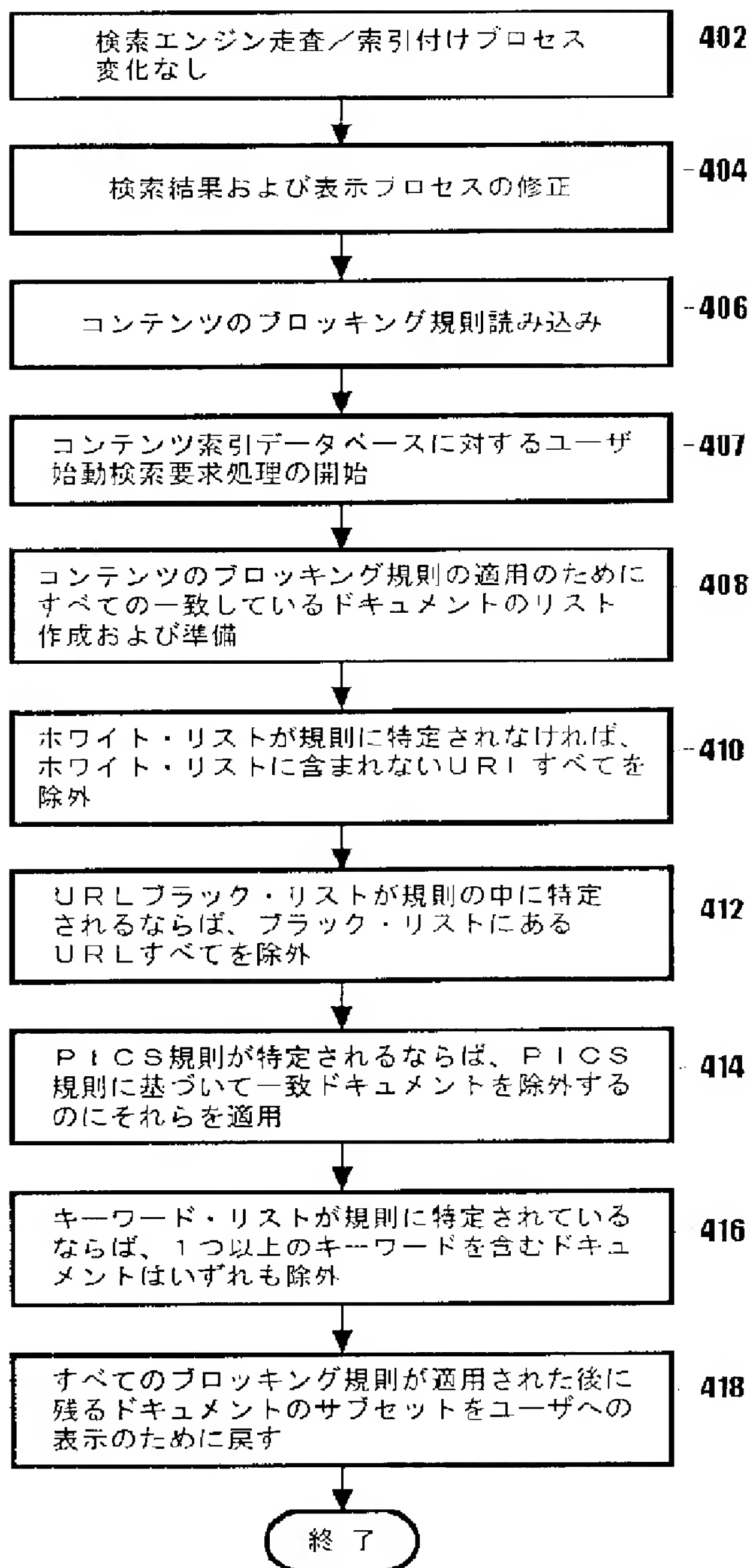
【図 3】

300



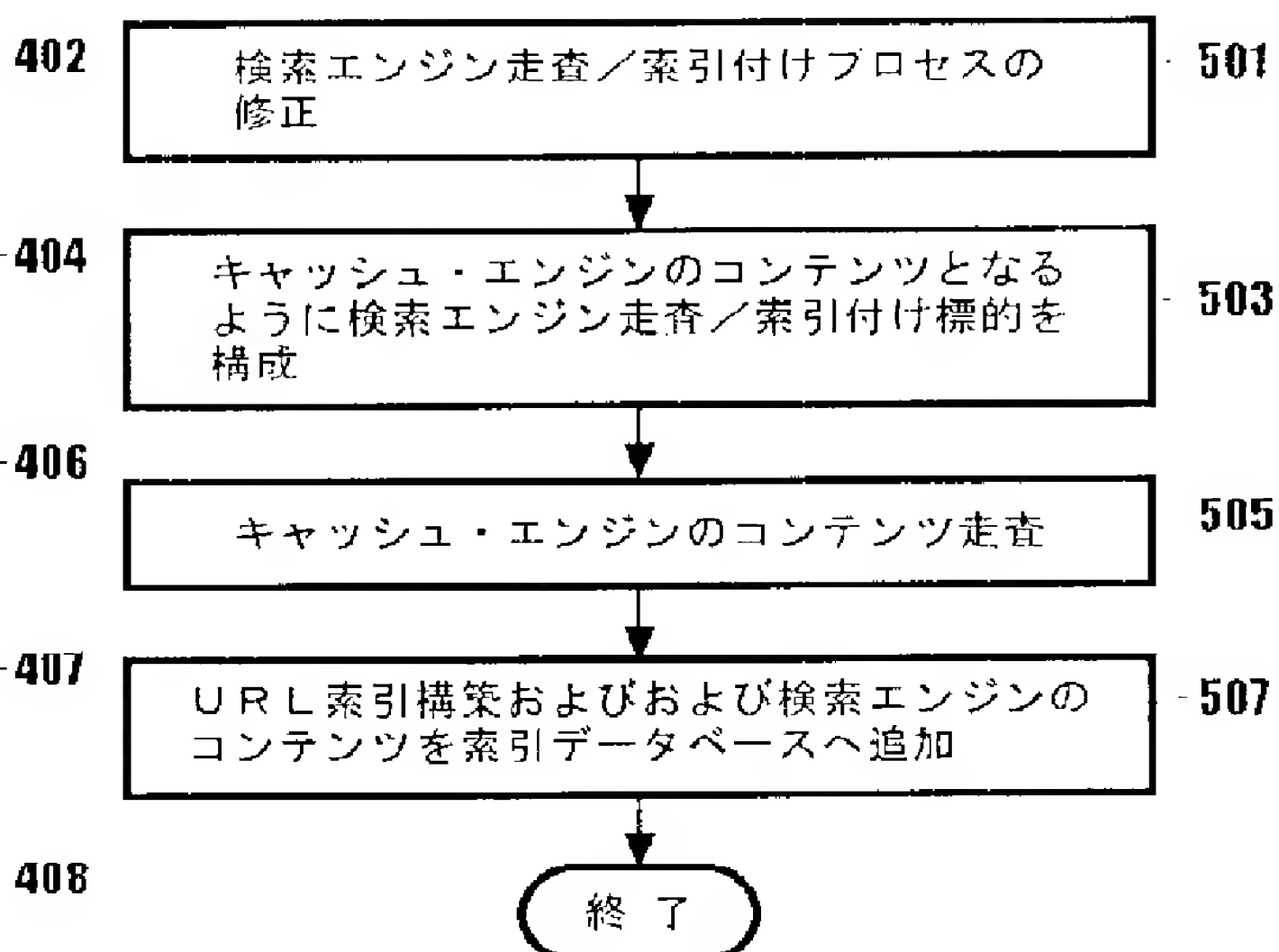
コンテンツ走査／索引付けフェーズの間の実行

【図4】

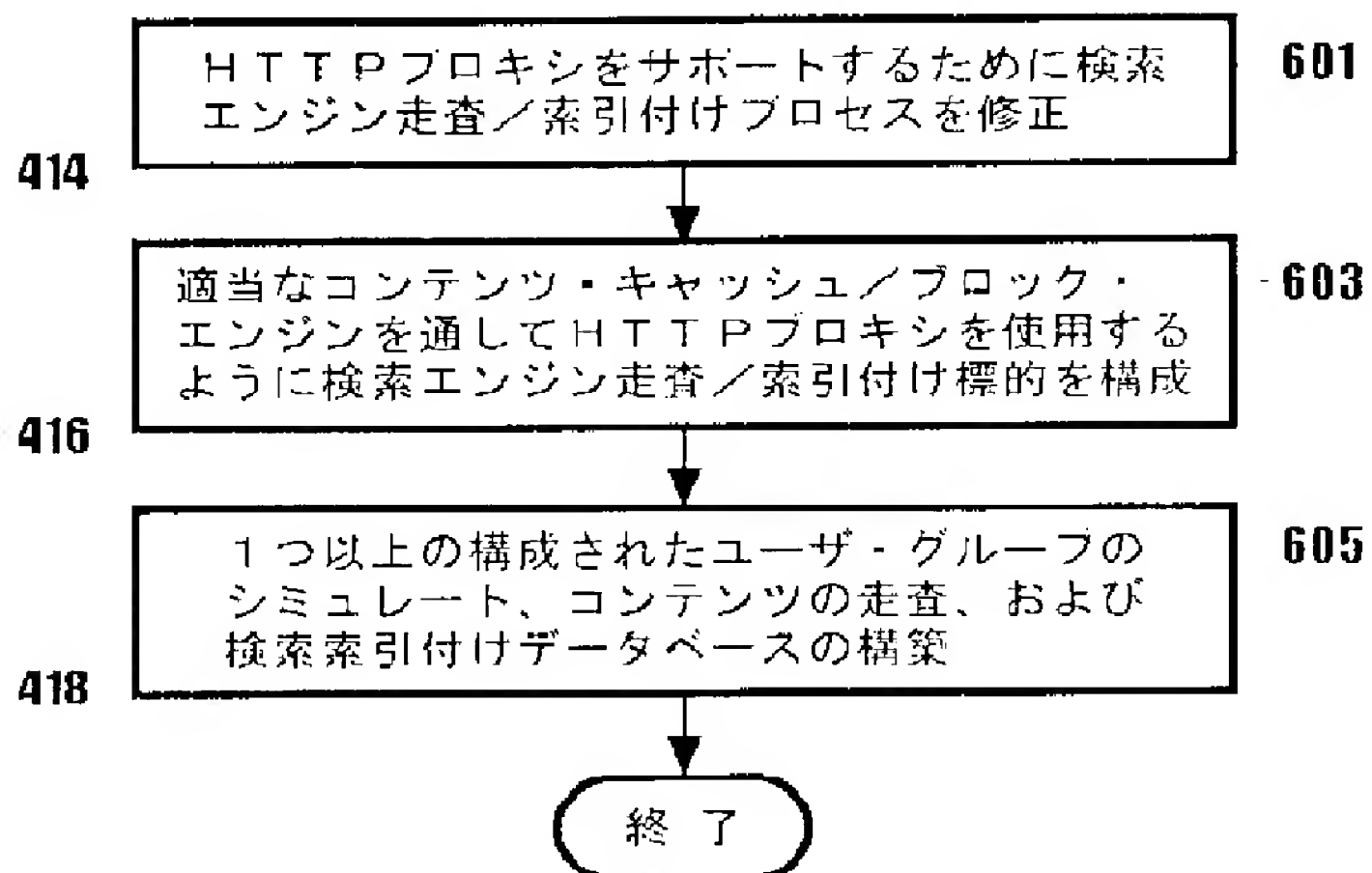
400

走査結果表示フェーズの間の実行

【図5】

500キャッシュ・エンジンのコンテンツから
索引付けデータベースの構築

【図6】

600キャッシュ・エンジンのコンテンツから
索引付けデータベースの構築

フロントページの続き

(51) Int. Cl. ⁷	識別記号	F I	ターミナル (参考)
----------------------------	------	-----	------------